Primer

Statistical learning: a powerful mechanism that operates by mere exposure

Richard N. Aslin*

How do infants learn so rapidly and with little apparent effort? In 1996, Saffran, Aslin, and Newport reported that 8-month-old human infants could learn the underlying temporal structure of a stream of speech syllables after only 2 min of passive listening. This demonstration of what was called statistical learning, involving no instruction, reinforcement, or feedback, led to dozens of confirmations of this powerful mechanism of implicit learning in a variety of modalities, domains, and species. These findings reveal that infants are not nearly as dependent on explicit forms of instruction as we might have assumed from studies of learning in which children or adults are taught facts such as math or problem solving skills. Instead, at least in some domains, infants soak up the information around them by mere exposure. Learning and development in these domains thus appear to occur automatically and with little active involvement by an instructor (parent or teacher). The details of this statistical learning mechanism are discussed, including how exposure to specific types of information can, under some circumstances, generalize to never-before-observed information, thereby enabling transfer of learning.

How to cite this article:
WIREs Cogn Sci 2017, 8:e1373. doi: 10.1002/wcs.1373

Whether one sees the newborn child as neurologically insufficient (Flechsig, 1920), cognitively confused (James, 1890), narcissistic (Freud, 1905), solipsistic (Piaget, 1927), or merely ugly (Hall, 1891), the distance between the new child and the walking, talking, socially discriminating, and perceptive person whom we see hardly 500 days later is awesome—*Kessen, Haith, and Salapatek*[1]

INTRODUCTION

Two giants in the field of animal learning created paradigms that illustrated how behavior can be shaped by stimuli that are paired with primary rewards such as food. Ivan Pavlov[2] used classical conditioning of the salivary response in dogs to a tone that signals impending delivery of meat powder, and B. F. Skinner[3] used operant conditioning in pigeons and rats to elicit pecking or bar-pressing behaviors when a cue signals the availability of a food pellet. In both of these paradigms, the associations among stimuli, responses, and rewards involve repeated exposure over many trials and a resultant slow time course of learning. How might such mechanisms enable the rapid learning that characterizes human (and non-human) development?

Two solutions were proposed. Edward Tolman[4] suggested that animals rarely rely on primary reinforcers to learn about their environment, but rather engage a mechanism akin to problem-solving in which the causal relations behind the associations observed by the animal are the fundamental motivation for and product of learning. Albert Bandura[5] suggested that learners gain as much information about the world by observing others as by actively engaging in the learning process themselves. There is

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: aslin@cvs.rochester.edu

Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY, USA

Conflict of interest: The author has declared no conflicts of interest for this article.

no question that *mature* learners, across a wide variety of species and domains, use both of these alternative mechanisms to speed up the learning process beyond that described by Pavlov and Skinner. But is it plausible for *naïve* learners, who are confronted with hundreds of potential stimuli each minute in their natural habitat, to determine which patterns of stimuli are worthy of their attention and likely to be informative for controlling their behavior without the guiding hand of a 'teacher' (i.e., rewards or punishments)? Research in the past 15 years has revealed that such a mechanism of learning 'by mere exposure' is both rapid and present in early development.

WHAT IS STATISTICAL LEARNING?

In both the lab and the natural environment, we learn about *distributions*. For the pigeon in a Skinner box, those distributions consist of which stimuli (e.g., a green light) are present immediately prior to behaviors (e.g., pecking a key) that are followed by a reward. There are many other stimuli that are present prior to the pecking response that do not lead to a reward. Thus, any animal that successfully learns the contingencies established by the experimenter has, in an implicit sense, created a 'list' of what works and what does not work in that particular context. Formulating such a list in the context of a Skinner box may seem fairly straightforward because the number of potential stimuli has been constrained and the relevant response has been shaped by the experimenter to simplify which aspects of the situation are relevant and how rapidly they are presented.

Now consider a more complicated situation in which there is no reward and the rate of stimulus presentation is greatly speeded up. Saffran, Aslin, and Newport[6] chose a task that must be solved by all children as they learn language: determining in a sample of speech spoken by their parents where one word ends and the next word begins. Words are composed of syllables from an inventory of sounds (the phonemes) used in the particular natural language to which the child is exposed. We know that children begin to speak their first words at about 12 months of age, so the process of solving the word-segmentation problem must be at least partially accomplished by the end of the first postnatal year. Moreover, what makes the word-segmentation problem particularly difficult is that the acoustic cues that define a word boundary are unreliable: some multi-word utterances have no acoustic cues to word boundaries ('Where are you?') and many words have obvious acoustic cues *within* themselves ('cookie'). If infants relied solely on these acoustic cues, they would incorrectly treat 'Where are you?' as one word and 'cookie' as two words.

Saffran et al.[6] asked whether infants could use distributional information to solve the word-segmentation problem. They eliminated all other cues that are present in natural speech, such as the pauses at the end of phrases and the modulations in pitch and duration that occur as syllables are put together to form fluent utterances (see Figure 1). It is important to note that the word-segmentation problem is not solved by the child's parents. Parents neither speak in one-word sentences (except for special words like 'no' or the child's name), nor do they define by explicit instruction what a word is. Rather, parents speak to their children in much the same way they speak to other adults, except at a somewhat slower rate and often with more modulation of pitch and amplitude. Thus, the naïve learner is confronted with an inventory of dozens of different syllables presented in child-directed speech at a rate of 3–5 per second, and must keep track of the sequential patterns embedded in a corpus of these sentences.

The task set up by Saffran et al. mimicked this situation by presenting 8-month-old infants with a continuous stream of speech syllables, without the benefit of pauses or intonation cues, and asked whether infants could 'compute' the distributions of how often each syllable occurred and how often combinations of syllables occurred. If infants could perform such computations, then in a post-familiarization test phase they should discriminate between sequences of syllables that varied in their statistical properties. In particular, it was hypothesized and confirmed in a follow-up experiment[7] that the probability of one syllable following another was a crucial source of information for solving the word-segmentation problem. The test phase presented infants with two types of trials: a triplet of syllables

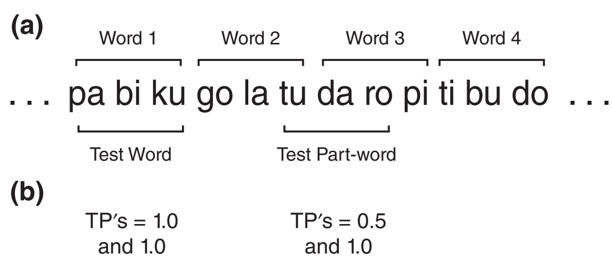

FIGURE 1 | Illustration of the stimuli used by Saffran et al.[6] to study statistical learning in 8-month-old infants. (a) The inventory of syllables and tri-syllabic words. (b) The statistical structure of the words and part-words. TP, transitional probability. See Supporting information for link to sound file.

with a perfectly predictable order (i.e., a word) and a triplet of syllables with a less predictable order because they spanned a word boundary (i.e., a part-word). Infants exhibited longer attention, as indexed by looking-time to a blinking light, to the less statistically consistent part-words than to the words. Note that keeping track of these distributions—the relative frequency of pairs of syllables and the transitional probability from one syllable to the next—when the syllables were presented at a rate of 4 per second and when no feedback is provided to guide the learning process, was considered to be highly implausible for an infant.

The results of Saffran et al.[6] and Aslin et al.[7] demonstrated that 8-month-old infants could, indeed, solve the word-segmentation problem only after a short exposure to a stream of speech whose words were defined solely on the basis of statistical cues. Importantly, as the stream of speech had never been heard by the infants prior to the experiment, and half of the infants were presented with a different stream that had the opposite set of statistical cues, yet all infants learned the relevant statistics, the overall pattern of results confirms that infants relied on statistical cues in these streams of speech to solve the word-segmentation problem in the lab. These results also demonstrate that months before infants have spoken their first word, and presumably before they know the meanings of all but a handful of words, they can utilize a statistical learning mechanism to extract those chunks of syllables that are most likely to become words in the language to which they are exposed (see Samuelson and McMurray, What does it take to learn a word?, *WIREs Cogn Sci*, also in the collection How We Develop). Finally, these results from infants confirm that statistical learning, like other implicit learning tasks,[8] is accomplished without direct evidence of conscious hypothesis testing or feedback from a teacher who identifies the problem to be solved or the means to solve it.

LANGUAGE-SPECIFIC OR DOMAIN-GENERAL?

One potential implication that could be drawn from the initial work on statistical learning is that it is a special mechanism employed by humans to rapidly acquire language. If that were true, then it would have limited relevance to the more general problem of enabling the naïve learner to rapidly acquire information in the environment that is relevant for solving any task. Two lines of evidence quickly refuted the language-specificity hypothesis. First, statistical learning operates over nonlinguistic stimuli including auditory tones,[9] visual shape-sequences,[10] and tactile patterns.[11] Second, statistical learning of speech streams is present in nonhuman species (rats) who will never acquire language.[12]

But statistical learning is not limited to the temporal domain. Fiser and Aslin[13] created a family of visual scenes (see Figure 2) that were composed from an inventory of visual shapes, analogous to the composition of words from speech syllables, but now in the spatial domain. By placing constraints on how the shapes were positioned in the scenes, the spatial statistics could be manipulated. At issue was whether passive viewing of a set of such multi-shape scenes, where each scene contained a subset of the shape inventory but always following the rules of shape combination, would enable adults to extract the shape combinations that were more predictable. A posttest after the 5-min exposure phase, in which 144 different six-shape scenes were presented every 2 seconds, revealed reliable sensitivity to the more

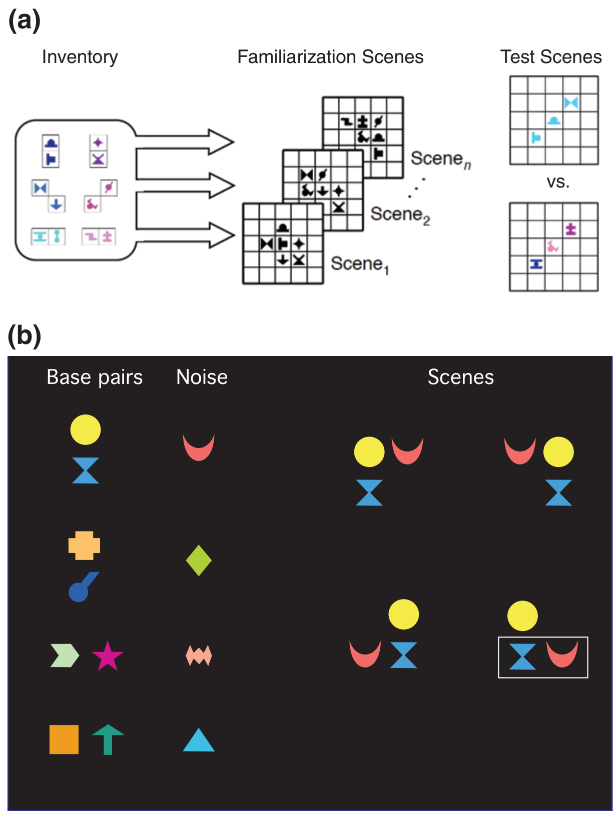

FIGURE 2 | The stimuli and design of the spatial version of the statistical learning task from Fiser and Aslin[13] with adults. (a) The inventory of shapes and a sample of their arrangement in 3 × 3 grids. The shape-pairs used during the post-test showing their joint and conditional probabilities. The stimuli and design of the analogous task used with 9-month-old infants by Fiser and Aslin.[14] (b) The inventory of shapes and sample scenes presented during familiarization. The shape pairs used during the post-test and their underlying statistics.

statistically coherent shape pairs. Further studies by Fiser and Aslin[15] showed that this sensitivity to spatial statistics was not limited to shape-pairs, but could be extended to whatever coherent structures (pairs, triplets, quadruplets) that were present in the family of scenes. Importantly, this spatial statistical learning ability is also present in 8-month-old infants.[14,16]

Coupled with the voluminous evidence of temporal statistical learning in humans for nonlinguistic stimuli as well as visual statistical learning in the spatial domain, it is now accepted that statistical learning is modality-, domain-, and species-general.

CONSTRAINTS ON STATISTICAL LEARNING

The presence of a powerful statistical learning 'engine,' especially early in development, provides a path toward rapid learning. But unless statistical learning is constrained, it suffers from being too powerful: learners do not have sufficient processing resources to extract and retain in memory the vast number of potential statistics that are available in the natural environment. Four types of constraints have been identified that enable statistical learning to be tractable given access to highly complex stimuli. (1) *Attention* that is manipulated by overt instruction to adults serves as a powerful 'filter' on what information is analyzed in a statistical learning task.[17,18] Although infants cannot be instructed in a similar manner, they are constrained by implicit attentional cues, such as the direction of gaze exhibited by a parent.[19] Moreover, Kidd, Piantadosi, and Aslin[20] showed that infants appear to have an implicit sense when attention is likely to lead to further information and when it is likely to be a 'dead end,' either because information is too simple (already known) or too complex (unknowable).

(2) *Perceptual biases* that enable some stimuli to be encoded more readily than others act as a powerful constraint. For example, stimuli that are presented adjacent in time are easier to associate than the same stimuli when separated by an intervening stimulus. Adults have greater difficulty learning nonadjacent statistics than adjacent statistics unless some common feature binds them together. In the musical domain, this binding is accomplished by having tones come from the same octave[21] and in the speech domain it is based on category similarity (e.g., vowels vs. consonants[22,23]).

(3) *Prosody* is the large-scale variation in pitch and duration that overlays the individual words in an utterance. Prosody operates in a language-specific way at the level of words by dictating patterns of strong and weak stress (e.g., *bag*gage in English vs. ba*guette* in French). But prosody also operates in a language-general way by partitioning sentences into phrases [e.g., (The little baby) (drank her milk)]. Shukla, White, and Aslin[24] showed that 6-month-old infants use language-general prosody, which signals a 'gap' between the two phrases, to prevent statistics from being computed across a phrase-boundary.

(4) *Primacy and Familiarity* also play a role in constraining statistical learning. Gebhart, Aslin, and Newport[25] showed that if adults are exposed to a single stream of syllables that had one statistical structure in the first half and a different structure in the second half, they learn the first but not the second. Even a change in the voice of the speaker is insufficient to induce adults to learn the second statistical structure unless that change in voice occurs repeatedly.[26] Familiarity also plays an important role in statistical learning. When sounds are unfamiliar, such as arbitrary noises rather than speech, adults are unable to encode them efficiently and statistical learning is prevented unless the rate of presentation is slowed considerably.[27]

MECHANISMS, MODELS, AND GENERALIZATION

The early work on statistical learning focused on extracting structured information from the input. But a hallmark of learning is the ability to go beyond the specific input to which the learner has been exposed. This process of generalization is crucial in many domains, like language, because a learner could never be exposed to every possible input. The ability to generalize to novel exemplars that bear some similarity to exemplars that have been experienced has been referred to as a process of abstraction or rule learning. Marcus, Vijayan, BandiRao, and Vishton[28] demonstrated rule learning in 9-month-old infants by exposing them to 16 different three-syllable sequences, separated by pauses, that each conformed to a uniform pattern (e.g., ABB, AAB, or ABA; see Figure 3). After familiarization, a new set of syllables was presented in a posttest, with half of the novel syllables forming the familiar pattern and half a novel pattern. Despite the fact that all post-test strings were novel, infants discriminated the novel pattern (e.g., a shift from AAB to ABB) from the familiar pattern.

As in the case of the original findings on statistical learning, these findings on rule learning could be interpreted as unique to language, but follow-up studies showed that this form of rule learning is

Syllable A \ Syllable B	di	je	li	we
le	le le di	le le je	le le li	le le we
wi	wi wi di	wi wi je	wi wi li	wi wi we
ji	ji ji di	ji ji je	ji ji li	ji ji we
de	de de di	de de je	de de li	de de we

FIGURE 3 | The design of the Marcus et al.[28] experiment on rule learning in which 9-month-olds were presented with a large inventory of syllables with a uniform ABB pattern. The test items presented after familiarization were composed of entirely novel syllables that either conformed to the familiar ABB pattern or exhibited a novel AAB pattern. The Gerken[29] experiment used two subsets of the overall inventory of stimuli from Marcus et al. Blue highlight = broad generalization. Red highlight = narrow generalization.

present in the visual modality,[30,31] for nonlinguistic auditory stimuli,[32] and in nonhumans.[33,34] Thus, rule learning is another example of a powerful mechanism that is both domain- and species-general.

Importantly, the categorical distinction between statistical learning and rule learning may be overemphasized. Gerken[29] showed that infants tested in a paradigm very similar to Marcus et al.[28] shifted their response from a broad generalization (e.g., AAB) to a narrower generalization (e.g., AA 'ends in a specific syllable') depending on how the learning strings were constructed (see Figure 3). Thus, infants switched from attending to 'abstract' patterning to attending to 'surface' features of the input, suggesting that rule learning and statistical learning may lie along a continuum rather than acting as discrete and separable mechanisms.

Further support for this hypothesis comes from Reeder, Newport, and Aslin[35] who conducted a series of experiments with adults and varied the patterning of how often certain nonsense words were paired with other nonsense words. The structure of this patterning defined the 'grammar' of sentences in this artificial language. In a passive exposure phase, adults listened to several hundred sentences containing the inventory of nonsense words, and in a subsequent test phase they rated whether the sentences were grammatical, according to the rules used to generate the sentences in the exposure phase. As expected, familiar sentences were rated as highly grammatical and novel sentences that violated the grammatical rules were rated as ungrammatical. But *novel* sentences that did not violate the grammatical rules were also rated as grammatical, thereby demonstrating generalization or transfer of knowledge from what adults heard in the exposure phase to novel sentences. Importantly, when the exposure phase presented sentences that had persistent 'gaps' in the way that certain words were not paired with other words, adults did not generalize to fill these gaps. For example, if the word 'dax' was followed by 'lif,' 'neem,' and 'zilk,' but the word 'flug' was only followed by 'lif' and 'neem,' the gap created by the absence of 'zilk' was judged as real and not accidental. Thus, depending on the patterning of words in sentences, adults shift the degree of generalization from an abstract rule (i.e., generalizing to novel sentences) to judging novel sentences as violating that rule (i.e., restricting generalization).

A critical issue in studies of statistical learning is the computational mechanism that enables extraction of structure and generalization to novel exemplars. It is possible that this mechanism is invariant over development, but it is also possible that new computational sub-mechanisms emerge as the overall statistical learning process expands to encompass more complex environmental input (see Oudeyer, What do we learn about development from baby robots?, *WIREs Cogn Sci*, also in the collection How We Develop). The original idea proposed by Saffran et al.[6] was that learners rely, at least in part, on computing transitional probabilities between adjacent syllables (see Figure 1). But a variety of other models have been proposed over the past 15 years, including ones that extract 'chunks' of syllables using a form of discrete sampling as attention waxes and wanes,[36,37] and ones that posit an implicit causal structure that generates the exemplars[38] (see Frank et al.[39] for a review of competing models). The causal structures approach has also been successfully applied to visual statistical learning in the spatial domain.[40] Importantly, a variety of models exhibit the gradient property of generalization from specific exemplars (statistical learning) to abstract principles (rule learning) that is observed in human learners,[28] including connectionist architectures[41] (see review by Aslin and Newport[42]).

WHAT DEVELOPS?

One of the most impressive aspects of statistical learning in humans is that it appears to be functional from birth.[43–45] Given such a powerful learning mechanism, it is seductive to conclude that there are no interesting developmental differences between the newborn and the adult. Although there is no compelling evidence for a *qualitative* developmental difference, there are certainly *quantitative* differences in how the statistical learning 'engine' gains access to structured information. For example, infants are notorious for having limited attentional skills, especially their ability to sustain attention.[46] Infants also have limited working memory,[47] have sluggish control over motor systems such as eye movements that sample the visual environment,[48] and must be exposed to certain types of stimuli before those stimuli are familiar enough to be easily encoded.[49] All these information-processing deficiencies limit the availability and/or quality of structured information upon which the statistical learning mechanism can operate. Thus, the development of statistical learning may be determined largely by the changes in 'input systems' that open a window to enable the extraction of higher-order structures that are initially 'invisible' because of low-level constraints. Future research on this issue will reveal how these quantitative efficiencies cascade to create qualitative changes in the organization of learning and memory systems over development.

ACKNOWLEDGMENTS

Preparation of this article was made possible, in part, by a research grant from NIH (HD-037082) to Richard Aslin and Elissa Newport.

REFERENCES

1. Kessen W, Haith MM, Salapatek PH. Human infancy: a bibliography and guide. In: Mussen PH, ed. *Carmichael's Manual of Child Psychology*, vol. 1. 3rd ed. New York: Wiley; 1970, 287–445.
2. Pavlov IP. *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex*. New York: Dover; 1927.
3. Skinner BF. *The Behavior of Organisms: An Experimental Analysis*. Appleton-Century: Oxford, UK; 1938.
4. Tolman EC. *Purposive Behavior in Animals and Men*. Berkeley, CA: University of California Press; 1932.
5. Bandura A. *Social Learning through Imitation*. Lincoln, NE: University of Nebraska Press; 1962.
6. Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. *Science* 1996, 274:1926–1928.
7. Aslin RN, Saffran JR, Newport EL. Computation of conditional probability statistics by 8-month-old infants. *Psychol Sci* 1998, 9:321–324.
8. Perruchet P, Pacton S. Implicit learning and statistical learning: one phenomenon, two approaches. *Trends Cogn Sci* 2006, 10:233–238.
9. Saffran JR, Johnson EK, Aslin RN, Newport EL. Statistical learning of tone sequences by human infants and adults. *Cognition* 1999, 70:27–52.
10. Kirkham NZ, Slemmer JA, Johnson SP. Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition* 2002, 83:B35–B42.
11. Conway CM, Christiansen MH. Modality-constrained statistical learning of tactile, visual, and auditory sequences. *J Exp Psychol Learn Mem Cogn* 2005, 31:24–39.
12. Toro JM, Trobalon JB. Statistical computations over a speech stream in a rodent. *Percept Psychophys* 2005, 67:867–875.
13. Fiser J, Aslin RN. Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychol Sci* 2001, 12:499–504.
14. Fiser J, Aslin RN. Statistical learning of new visual feature combinations by infants. *Proc Natl Acad Sci U S A* 2002, 99:15822–15826.
15. Fiser J, Aslin RN. Encoding multielement scenes: statistical learning of visual feature hierarchies. *J Exp Psychol Gen* 2005, 134:521–537.
16. Wu R, Gopnik A, Richardson DC, Kirkham NZ. Infants learn about objects from statistics and people. *Dev Psychol* 2011, 47:1220–1229.
17. Toro JM, Sinnett S, Soto-Faraco S. Speech segmentation by statistical learning depends on attention. *Cognition* 2005, 97:B25–B34.
18. Turk-Browne NB, Junge J, Scholl BJ. The automaticity of visual statistical learning. *J Exp Psychol Gen* 2005, 134:552–564.
19. Wu R, Kirkham NZ. No two cues are alike: depth of learning during infancy is dependent on what orients attention. *J Exp Child Psychol* 2010, 107:118–136.
20. Kidd C, Piantadosi ST, Aslin RN. The Goldilocks effect: human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS One* 2012, 7:e36399.
21. Creel SC, Newport EL, Aslin RN. Distant melodies: statistical learning of nonadjacent dependencies in tone sequences. *J Exp Psychol Learn Mem Cogn* 2004, 30:1119–1130.
22. Bonatti LL, Pena M, Nespor M, Mehler J. Linguistic constraints on statistical computations: the role of consonants and vowels in continuous speech processing. *Psychol Sci* 2005, 16:451–459.
23. Newport EL, Aslin RN. Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cogn Psychol* 2004, 48:127–162.

24. Shukla M, White KS, Aslin RN. Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proc Natl Acad Sci U S A* 2011, 108:6038–6043.
25. Gebhart AL, Aslin RN, Newport EL. Changing structures in midstream: learning along the statistical garden path. *Cogn Sci* 2009, 33:1087–1116.
26. Weiss DJ, Gerfen C, Mitchel AD. Speech segmentation in a simulated bilingual environment: a challenge for statistical learning? *Lang Learn Dev* 2009, 5:30–49.
27. Gebhart AL, Newport EL, Aslin RN. Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychon Bull Rev* 2009, 16:486–490.
28. Marcus GF, Vijayan S, Bandi Rao S, Vishton PM. Rule learning by seven-month-old infants. *Science* 1999, 283:77–80.
29. Gerken L. Decisions, decisions: infant language learning when multiple generalizations are possible. *Cognition* 2006, 98:B67–B74.
30. Johnson SP, Fernandas KJ, Frank MC, Kirkham N, Marcus G, Rabagliati H, Slemmer JA. Abstract rule learning for visual sequences in 8- and 11-month-olds. *Infancy* 2009, 14:2–18.
31. Saffran JR, Pollak SD, Seibel RL, Shkolnik A. Dog is a dog is a dog: infant rule learning is not specific to language. *Cognition* 2007, 105:669–680.
32. Marcus GF, Fernandes KJ, Johnson SP. Infant rule learning facilitated by speech. *Psychol Sci* 2007, 18:387–391.
33. de la Mora DM, Toro JM. Rule learning over consonants and vowels in a non-human animal. *Cognition* 2013, 126:307–312.
34. Murphy RA, Mondragon E, Murphy VA. Rule learning by rats. *Science* 2008, 319:1849–1851.
35. Reeder PA, Newport EL, Aslin RN. From shared contexts to syntactic categories: the role of distributional information in learning linguistic form-classes. *Cogn Psychol* 2013, 66:30–54.
36. French RM, Addyman C, Mareschal D. TRACX: a recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychol Rev* 2011, 118:614–636.
37. Perruchet P, Vinter A. PARSER: a model for word segmentation. *J Mem Lang* 1998, 39:246–263.
38. Goldwater S, Griffiths TL, Johnson M. A Bayesian framework for word segmentation: exploring the effects of context. *Cognition* 2009, 112:21–54.
39. Frank MC, Goldwater S, Griffiths TL, Tenenbaum JB. Modeling human performance in statistical word segmentation. *Cognition* 2010, 117:107–125.
40. Orban G, Fiser J, Aslin RN, Lengyel M. Bayesian learning of visual chunks by human observers. *Proc Natl Acad Sci U S A* 2008, 105:2745–2750.
41. Seidenberg MS, Elman JL. Do infants learn grammar with algebra or statistics? *Science* 1999, 284:434–435.
42. Aslin RN, Newport EL. Statistical learning: From acquiring specific items to forming general rules. *Curr Dir Psychol Sci* 2012, 21:170–176.
43. Bulf H, Johnson SP, Valenza E. Visual statistical learning in the newborn infant. *Cognition* 2011, 121:127–132.
44. Gervain J, Macagno F, Cogoi S, Pena M, Mehler J. The neonate brain detects speech structure. *Proc Natl Acad Sci U S A* 2008, 105:14222–14227.
45. Teinonen T, Fellman V, Naatanen R, Alku P, Huotilainen M. Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neurosci* 2009, 10:21.
46. Richards JE. Attention in young infants: a developmental psychophysiological perspective. In: Nelson CA, Luciana M, eds. *Handbook of Developmental Cognitive Neuroscience*. Cambridge, MA: MIT Press; 2008.
47. Blaser E, Kaldy Z. Infants get five stars on iconic memory tests: a partial-report test of 6-month-old infants' iconic memory capacity. *Psychol Sci* 2010, 21:1643–1645.
48. von Hofsten C, Kochukhova O, Rosander K. Predictive tracking over occlusions by 4-month-old infants. *Dev Sci* 2007, 10:625–640.
49. Kelly DJ, Liu S, Lee K, Quinn PC, Pascalis O, Slater AM, Ge L. Development of the other-race effect during infancy: evidence toward universality? *J Exp Child Psychol* 2009, 104:105–114.

